图书情报工作
第 64 卷 第 2 期 2020 年 1 月

大数据环境下应急信息质量评估体系研究*

■ 徐文强 刘春年 周涛
南昌大学管理学院 南昌 330031

摘 要:[目的/意义]针对当前大数据时代下应急信息质量管理面临的新挑战,提出大数据环境下应急信息质量评估体系,以帮助筛选出高质量的应急信息。[方法/过程]基于大数据环境下应急信息的特征以及在应急管理中对于高质量应急信息的需求分析,设计并构建大数据环境下的应急信息质量评估体系;并且利用云模型对 EPS 数据库和 EM - DAT 突发事件数据库中国内的应急信息进行案例研究。[结果/结论]评估结果与实际情况较为相符,说明该评估体系对指导应急信息质量评估工作具有一定作用。

关键词:大数据 应急信息 信息质量 评估体系
分类号:G203
DOI:10.13266/j.issn.0252 - 3116.2020.02.006

应急管理关乎民生大计,高质量的应急信息能为应急管理提供信息支撑。但大数据环境下,应急信息的质量存在着众多的挑战:信息噪声的产生为应急信息管理带来了干扰,严重影响着应急信息的准确性、科学性;多源异构的信息数据为应急信息的管理带来了困扰以及动态性、及时性的要求;信息量的爆炸式发展导致了信息的冗余、重复等。大数据时代的到来使信息环境发生了改变,信息环境的改变对应急信息质量提出了新的要求,应急管理活动也对应急信息质量提出了实践需求。应急信息质量评估是检测其质量是否合格的关键。应急信息质量评估的目的在于剔除低质量的、错误的信息,筛选出高质量的、符合应急决策要求的应急信息,并为应急信息质量的进一步提升与治理提供可靠依据。经过筛选、质量有所提升的应急信息能够更好地辅助应急管理工作。此外,大数据质量与信息质量管理与评估也是当今研究的热点,因此有必要对大数据环境下的应急信息质量评估进行探讨。

1 文献综述

在信息质量评估的相关研究中,查先进等[1]指出了信息资源质量的重要性,并且阐述了信息资源质量的内涵和外延,构建了 4 维度 16 指标的信息资源质量评估指标体系,对信息资源质量评估的定性、定量、半定性方法进行了对比研究。宋维翔等[2]指出信息质量的指标中应包括权威性、原创性、适量性等。王仙雅等[3]指出信息质量的指标对感知有用性和持续搜寻行为的影响主要可分为 4 类。王素芳等[4]对高校图书馆信息共享服务质量进行评估,构建了包括“实体空间”“信息控制”和“服务效果”3 个维度、21 个指标的评估指标体系。

在应急信息质量研究方面,相关研究多为对应急信息质量某一指标和应急信息服务质量的探讨,其中朱益平等[5]提出了应急信息的数据准确性测量框架。刘春年等[6]以应急网站为研究对象,对应急网站的信息服务质量进行了评价及差异发现。袁维海[7]指出应急信息质量评价在应急信息筛选中发挥着重要的作用,并且构建了应急信息质量评价指标体系。刘春年等[8]在研究中发现缺乏评估应急产品信息可靠性的技术是寻求应急产品信息的一个限制。已有学者从不同角度对大数据环境下的应急信息进行探索,且研究多集中于如何对其进行融合利用以支持决策、实现科学管理。郭路生等[9-10]探索了大数据环境下的应急情报需求的开发模式和要素,并利用企业架构(enterprise architecture)对大数据环境下的应急信息资源进行规划,实现了应急信息在不同部门、不同系统之间的共享、系统集成。郭春侠等[11]针对大数据背景下应急信

* 本文系国家自然科学基金项目“大数据环境下应急信息质量与信源可信度双路径作用机理研究”(项目编号:71663038)研究成果之一。
作者简介:徐文强(ORCID:0000 - 0001 - 5919 - 5904),硕士研究生;刘春年(ORCID:0000 - 0002 - 8764 - 134X),教授、博士生导师,E-mail:pan81706@163.com;周涛(ORCID:0000 - 0002 - 7282 - 9462),本科生。
收稿日期:2019 - 06 - 04 修回日期:2019 - 08 - 31 本文起止页码:50 - 58 本文责任编辑:易飞

息管理需求探讨了应急管理人才培养。李露[12]从多角度指出大数据环境下要实现对应急信息的多方面、多层次的融合。

目前国内外对应急信息质量评估理论的研究较为薄弱,应急信息质量评估理论尚需进一步完善。其中袁维海[7]认为应急信息质量评估是在应急信息管理过程中对应急信息质量进行筛选的一环,是对应急信息管理各阶段进行信息筛选的关键。H. Seppänen 等[13]认为及时性、准确性是应急信息质量的重要反映。学者 A. C. M. Oliveira 等[14]提出了以情境评估(scenario assessment)的方法对应急信息的质量进行评估,这种评估方案为决策者提供了在不确定的环境下理解突发事件情况的模型。学者 R. B. Street 等[15]总结了灾害信息管理存在的主要挑战以及信息质量控制机制与应急信息的可理解性,倡导人们加强对信息质量在灾害防护及应对中作用的重视。在对应急信息质量评估方法的选取上,大都沿用信息质量的评估方法。在评估标准上并未形成统一的质量标准。

综上,笔者通过对现有研究结果的分析,发现目前应急信息质量研究多集中于对其服务质量的研究以及对某个指标的具体探讨,但是还应加强更深层次上对应急信息质量实质的探讨。目前信息质量评估指标体系研究所构建的指标体系,大都借鉴信息质量评估的指标,对应急信息特性的分析与应用尚需深入。同时,在大数据时代,应急信息环境、应急信息特征、应急信息管理工作都发生着改变,针对应急信息所表现出的复杂维度、海量、结构复杂等特性,都有必要进行更深入的研究,因此亟需建立符合大数据环境特点的应急信息质量评估指标体系。

基于以上发现,本文在研究相关理论与实践的基础上,结合大数据环境下应急信息的特征与实践需求,提出 3 维度 8 指标的大数据环境下应急信息质量评估体系,而后本文选取具有代表性的 EPS 数据平台与 EM-DAT 突发事件数据库,随机选取其中 9 类应急信息,利用评估体系与云模型从不同角度对两个数据库进行评估分析,为提升其应急信息质量提供评估依据。

2 大数据环境下应急信息质量评估理论阐述及实践需求

2.1 大数据环境下应急信息质量评估理论阐释

2.1.1 信息质量评估理论

信息质量是信息质量管理的关键内容,对信息资源质量控制发挥着十分重要的作用。信息质量评估标准是评判信息质量的“标尺”,W. L. Yang 等[16]从信息来源的全面与稳固性两方面,提出对信源的自动评估方法。S. Madnick 等[17]从信息的可再生性与可理解性角度对信息进行评估。J. M. Juran[18]从信息的需求出发,指出信息质量评估标准应是根据用户信息使用要求的满足程度制定。B. K. Kahn 等[19]指出信息质量评估标准的定义应从产品质量与服务质量两方面出发。综合以上学者研究,对信息质量标准的定义应遵循准确、完整、系统全面的原则,同时考虑使用要求与效用两方面,对其进行评估。

2.1.2 应急信息质量内涵界定

由于应急信息服务于应急工作,因此其质量评估时要充分考虑其特殊性。在对应急信息质量进行评估时,我们应对应急信息的内涵进行探讨,有学者认为应急信息可分为两类:狭义上指灾害信息、灾害信息资源,广义上指一切与应急管理相关的信息[20]。也有学者认为应急信息是所有为应急决策提供支持作用的信息(如灾情现场信息、预案信息、预警信息等)。本文依据当前学者研究成果,同时结合应急管理流程,将应急信息定义为应急过程中产生的与应急相关的信息,具体而言可分为如下几类:

(1)应急预防与准备阶段:应急教育信息(自救、互救等基本常识信息),应急预案信息(应急准备,应急协调与保障(包括物资、通信、避难场所等信息)),应急政策法规(应急规范、法律法规等),应急专家信息等。

(2)突发事件预警与监测阶段:灾害预警监测信息,危害风险信息,应急信息报送,以及相关预警措施等。

(3)应急处置与救援阶段:新闻媒体报道信息,网络舆情,应急处置中产生的有关人财物等信息,救援与处置措施,应急管理过程共享信息。

(4)恢复重建阶段:灾害损失评估,应急事件处置报告,灾后政府发布声明,灾后重建规划等[21]。

由应急信息内涵可以总结应急信息资源的特征:①来源广。应急信息产生自应急管理各环节的不同主体,加大了对应急信息质量进行评估的难度。②信息量庞大。应急信息信源广泛的特性决定了信息量存在短时间爆发的风险,而且在大数据环境背景下这一特性尤为突出。③不确定性。互联网、移动通信、物联网技术的广泛使用,加快了信息传播速度,信息流通速度的加快使得在传播过程中难免受信息噪声干扰,从而不确定性增大。④时效性强。信息的价值随时间呈递减态势,应急信息尤为明显。一旦错过最佳传递时间,

应急信息的价值及有用性都将大大降低[20]。

2.1.3 云理论和云模型

1995 年,学者李德毅首次提出了云理论。而为了实现将自然语言值所表现的定性概念与定量数据之间的不确定性转换,学者建立了云模型。云由许多云滴构成,云滴是定性概念向定量值间的映射,其整体形状是对定性概念的刻画。云的特征由 3 个数值描述:期望 *Ex*,熵 *En*,超熵 *He*。其中:期望 *Ex* 最能表现定性概念;熵 *En* 表示对定性概念刻画的模糊程度,熵值越大,所表现的定性概念越模糊;超熵 *He* 则表示云滴的离散程度,超熵越大,云就越“厚”。自云理论提出至今,云模型已成功地运用到了自然语言处理、数据挖掘、决策分析与智能控制等领域。

现实世界中,要使事物的性质更加真实、可靠,需要用自然语言进一步描述。大数据环境下的应急信息评估体系,指标体系中的变量不可避免地要受到各种不确定性因素的综合影响,不确定性描述的最好方法是自然语言,云模型可以通过大量的定量概念值及其确定度来实现对定性概念的刻画。以概率论和模糊理论为基础的云模型,能够将决策对象的模糊性与随机性很好地结合起来,实现定性概念与定量值之间的自然映射,最大程度地减少映射过程造成的信息丢失[22]。

2.2 大数据环境下应急信息质量评估需求分析

2.2.1 大数据环境下应急信息资源环境及特点分析

就大数据环境下的应急信息特征而言,物联网、互联网技术、计算技术广泛运用于应急管理领域,使得应急信息的来源变得广泛,这些应急信源时刻产生着预警信息、地理信息、交互信息等,这些信息体现出体量庞大(volume)的特点;并且这些应急信息中既有图片、文字、地理位置,又有网络社交舆情信息等结构化非结构化的信息,使应急信息呈现出结构多样(variety)的特征;由于应急事件的突发性、灾情舆情的动态性以及网络技术被运用到应急管理中使应急信息表现的传播迅速性,应急信息呈现高速性(velocity);另外由于应急信息数据体量大、信息的价值有限,使其呈现价值稀疏性(value);此外由于应急信息在应急管理中的重要地位,使得对其质量有较高的要求,表现出对其真实性(veracity)的追求。这些都符合大数据的特征,因此不能再用传统的手段与思维来处理大数据环境下的应急信息。大数据环境下对应急信息质量的要求如表 1 所示:

表 1 大数据环境下对应急信息质量的要求

应急信息来源	应急信息内容	应急信息特征	应急信息质量要求
物理监测传感器	突发事件监测、预警,灾情环境、时间、地点,交通信息等	信源广、结构复杂、信息噪声干扰	准确、有效
网络传感器	网络舆情,灾情描述等	来源、形式多样,价值密度低,错误信息、干扰信息	准确、真实
应急信息报道	新闻报道,政府公告	准确、时效性追求	准确、及时
应急信息库	预案、历史突发事件信息、救援案例、组织机构信息等	冗余、标准规范不一	准确、衔接性、一致性

2.2.2 大数据环境下应急信息资源质量评估需求

大数据环境为应急信息提出了新的质量要求,作为应急信息质量管理的重要环节,大数据应急信息质量评估在信息质量管理中的地位十分重要。因此对应急信息质量评估的理论、方法、工具、标准的建立与创新都迫在眉睫。通过评估应急信息质量,从而进行下一步对应急信息质量的筛选,使应急信息质量在技术上和管理上均有所提升,才能提高应急信息管理水平,为应急决策提供准确、及时、高质量的应急信息资源。大数据环境下应急信息质量评估的需求如图 1 所示:

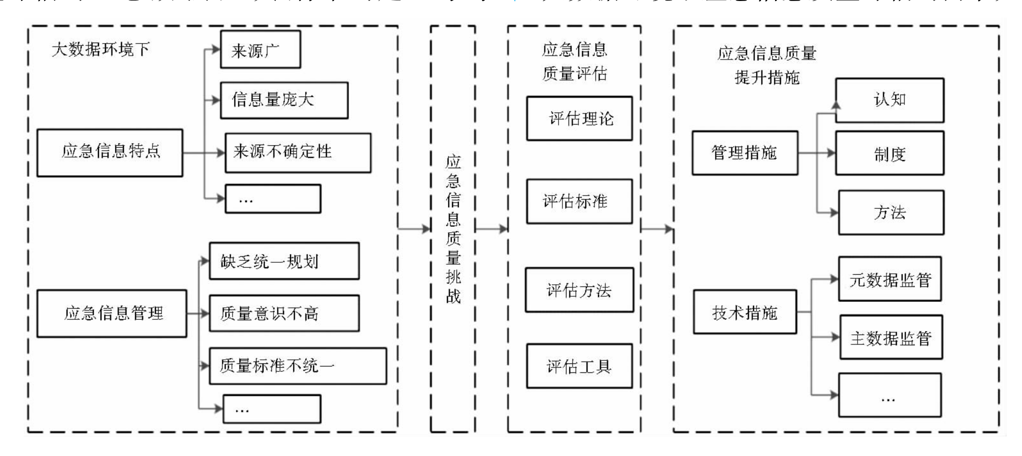


图 1 大数据应急信息评估需求

3　大数据环境下应急信息质量评估体系构建

3.1　大数据环境下应急信息质量评估指标体系概要设计

目前,基于对信息质量的不同理解,专家学者从不同角度提出了各自的评估指标体系[23－25]。但是由于缺乏坚实的信息质量理论,对信息质量评估的研究还处于探索阶段,缺乏一致认同的评估体系。处于大数据环境下的信息资源表现出复杂的特征,且其质量是一个多维的概念,因此对其进行评估的维度与指标都存在复杂多样的特征。根据现有研究不难发现大数据环境下信息质量的分析应包含其信息的内容,同时还应对其效用、表达、来源等进行分析。同时大数据环境下的信息质量应从信息角度的广泛维度以及其自身特点的独特维度进行考虑,既包括独立于任务的指标也包括其依赖任务的指标。在选取信息质量指标时既要考虑到其作为信息的普通性质,又要考虑其作为特殊任务的独特性质。

总结国内外学者对同类型的信息质量维度的定义不难发现,学者大都从 3 个角度对其进行探讨:①内容角度. 信息的准确性是大数据环境下的追求,大数据环境下,信息噪声以及复杂、多形式的信息使得准确性成为应急信息质量的要素。另一方面,大数据环境下人们对待信息的理念与方式发生了改变,人们不再追求信息的绝对精确,而是追求其内在的联系,因此误差性也是信息质量的影响因素;应急信息对于时效性的追求以及大数据环境下随时间衰变的信息价值使信息的及时性成为信息质量的要素;大数据的重要表现即根据不同情境满足用户不同信息需求,因此应急信息对于用户需求的满足程度,即应急信息的适用性,也影响着应急信息质量。②信息使用者角度。通过对信息进行描述,信息的使用者才能更好理解并利用信息,因此信息必须是可信的。在应急管理各环节、不同部门间、不同数据库之间、数据库内部均存在着规范与标准不统一、信息追踪困难、对信息进行前后联系困难、前后不可衔接或不一致等现象,使得应急信息质量存在问题。③信息的价值角度。应急信息具有信息的普通性质,信息所处的制度环境、信息能否被获得、信息的价值以及不同信息所承载的内容是否相关等都是其重要质量要素。

本文在对比国内外大数据质量评估指标、信息质量评估指标、应急信息质量评估指标的基础上,参考并总结出相关研究领域专家的研究成果,初步选取了影响大数据环境下应急信息质量评估的 3 个维度的 12 个指标:信息内容维度包含准确性、及时性、适用性、误差性;从信息使用者角度延伸出的信息描述维度包含可信性、一致性、可解释性、可衔接性;从信息价值角度延伸出的信息约束维度分为可获得性、有效性、相关性、制度环境。

3.2　专家咨询

经典的德尔菲法在确立初步指标时并不用给出指标的备选方案,而是通过向专家咨询后得出初步的指标方案。本文依据前文对同类指标的研究与实践,分析大数据应急信息质量的影响因素,已初步确立了指标方案。所使用的德尔菲法步骤如下:

3.2.1　确定咨询专家

经过多方联系确定参加本次调查的专家共有 15 名,这 15 名专家所处领域分别为应急教育领域、应急研究领域、应急实践工作领域,且具有专业知识和丰富实践经验。

3.2.2　设计调查问卷

在选取大数据环境下应急信息质量评估的指标时,根据上文已对国内外研究结论进行过初步筛选,此次是以向专家发放问卷的形式进行调研评估。此次问卷调查内容主要包括研究目的阐述、研究背景介绍、德尔菲法流程介绍、先拟定评估的具体指标。

3.2.3　开展专家咨询

(1)第 1 轮专家咨询:首先给出确定的评估指标,每位专家根据自身知识和经验对每个指标的影响程度评分,影响程度得分划分为 LiKert 型标度,采用 1－5 分进行标记,具体是“不大”“一般”“大”“很大”“极大”。最终对专家打分结果进行汇总,统计分析得到最终专家对备选指标的打分情况。

(2)第 2 轮专家咨询:依据第 1 轮专家咨询统计分析的结果,每个专家对每一组指标的判断结果的一致意见[18－19](2/3 专家对指标的判断结果 3)、平均值的分布情况向专家进行汇报,第 2 轮专家根据第一次的结果,重新对指标进行评价。在德尔菲法中,专家的意见都是从不一致到一致的探讨过程。本文的研究中,因为专家在第 2 轮的意见基本趋于一致,所以本文只进行了 2 轮专家咨询。专家咨询意见的协调性计算结果情况如表 2 所示:

表 2　2 轮专家咨询协调性情况

计算结果	第 1 轮	第 2 轮
协调系数 W	0.325 6	0.691 2
χ_1^2	3.581 2	10.368 0

(3)第 3 轮专家咨询:为了得出指标体系的权重,发放问卷由专家对指标间的重要性进行比较,如表 3 所示:

表 3 判断矩阵标度及其含义

序号	重要性等级	C_{ij}赋值
1	i,j 两元素同等重要	1
2	i 元素比 j 元素稍重要	3
3	i 元素比 j 元素明显重要	5
4	i 元素比 j 元素强烈重要	7
5	i 元素比 j 元素极端重要	9
6	相邻判断值的中间值	2、4、6、8

(4)调查结果处理:对专家咨询的结果进行计算处理。

3.3 大数据环境下应急信息质量评估指标体系的确定

挑选出第二轮调查形成的专家组一致意见大于等于 3 的指标作为评估指标,按照此标准通过对 12 个指标的 2 轮专家咨询,剔除了误差性、可信性、相关性和制度环境指标,共筛选出 8 个指标。其中在大数据环境下对信息质量的追求不再是精确而是追求一定的准确性,因此对于误差性在准确性中有一定的体现,而相关性在一致性和可衔接性中也有一定体现,应急信息的质量对于所存在的制度环境影响不大,最终确定了大数据环境下应急信息质量评估指标体系,如图 2 所示:

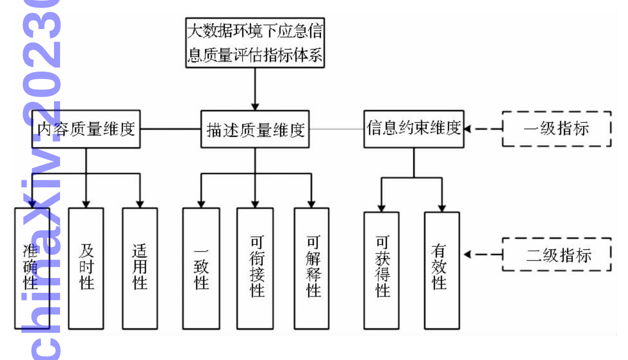


图 2 大数据环境下应急信息质量评价指标体系

3.3.1 内容质量维度

信息是内容的重要载体,内容是信息最基本的特征。在内容质量内涵上主要包括信息的准确性、及时性、适用性:①准确性是信息质量评估的重要指标之一,由于大数据环境下信息噪声的存在,低质量、低价值、错误和干扰信息充斥其中,对信息的处理、决策都产生了挑战甚至是干扰,准确性是对现实情况的真实反映,准确性越高反映真实情况程度越高。②及时性表现为信息收集、传输、决策、发布的时效、动态以及时间跨度。自然灾害和突发事件瞬息万变,而大数据环境下信息的"1 秒"定律也反映着时效性对价值的影响,因此对于及时性,应急管理决策者有着极其重要的要求。③适用性表现为信息是否满足用户、决策者的需求,大数据环境下的信息质量对能否满足信息用户的需求也有着极高的要求,因此在对应急信息进行收集、整理、数据挖掘等过程中应当充分考虑用户和决策者的需求。

3.3.2 描述质量维度

信息的描述质量表现为在信息语义描述状态下的质量,描述质量具体含义表现为信息的一致性、可衔接性和可解释性:①一致性表现为在数据库内逻辑的一致性以及不同数据库间的一致性;②可衔接性指所呈现的信息目录、分类及相关链接等是否在同一应急部门、不同部门上下可衔接、是否相关;③可解释性指应急信息便于决策、用户正确理解和使用的程度。

3.3.3 信息约束维度

应急信息质量进行评估时,需考虑信息的内容质量、描述质量外还需对其作为普通信息的质量约束标准,这也是应急信息非常重要的质量特征,具体包括:①应急信息是否易于用户、决策者获得,是应急信息效用的重要表现。②应急信息的有效性也是信息价值的重要表现,因此应急信息在应急管理活动中对效率的影响也是衡量其信息质量的重要指标。

3.4 确定属性权重

通过第 3 轮专家咨询建议确定指标权重,由于在指标体系的构建过程中,专家对各指标的内容与重要性有所了解,因此本文利用层次分析法对指标权重进行确定,其中各指标的重要性比较见表 3。根据专家打分情况构建相关判断矩阵,并对判断矩阵进行一致性判断,最终得出一级、二级指标的权重情况,如表 4 所示:

表 4 大数据环境下应急信息质量评估指标体系及权重

目标层	一级指标/权重	二级指标/权重/综合权重
大数据环境下应急信息质量	内容质量/0.412 6	准确性/0.412 6/0.170 2
		及时性/0.327 5/0.135 1
		适用性/0.259 9/0.107 2
	描述质量/0.259 9	一致性/0.4/0.104 0
		可衔接性/0.2/0.052
		可解释性/0.4/0.103 9
	信息约束/0.327 5	可获得性/0.5/0.163 8
		有效性/0.5/0.163 8

4 基于云模型的大数据环境下应急信息质量评估案例研究

由于信息环境的改变,应急信息所表现的特征完全符合大数据所表现的"4V"特征,是一种典型的大数据[28]。因此可以利用大数据相关技术对其进行处理。而目前对大数据环境下应急信息质量评估的方法还需要不断丰富。云理论自 1995 年提出以来,经过学者的深

入应用研究，已经广泛地运用在信息质量的评估、自然语言处理以及数据挖掘、智能决策等领域，理论与应用都较为成熟。因此本文尝试将云模型引入大数据环境下应急信息质量评估中，充分考虑大数据环境下应急信息的模糊性和不确定性，建立应急信息质量评价因素影响强度的定性定量转换，该模型为大数据环境下应急信息评估、管理与数据挖掘、智能决策提供了新的选择。

4.1 基于云模型的评估流程设计

本文尝试将云模型引入大数据环境下应急信息质量评估中。设计评估步骤流程如下：

(1)评估指标集；

(2)将不确定语言形式的评估信息转换为云，即设立应急信息质量评估的等级标准；

(3)引入属性权重；

(4)构建专家对各个案例评估的多层次综合云；

(5)得出评估结果。

基于云模型的大数据环境下应急信息质量评估具体流程如图3所示：

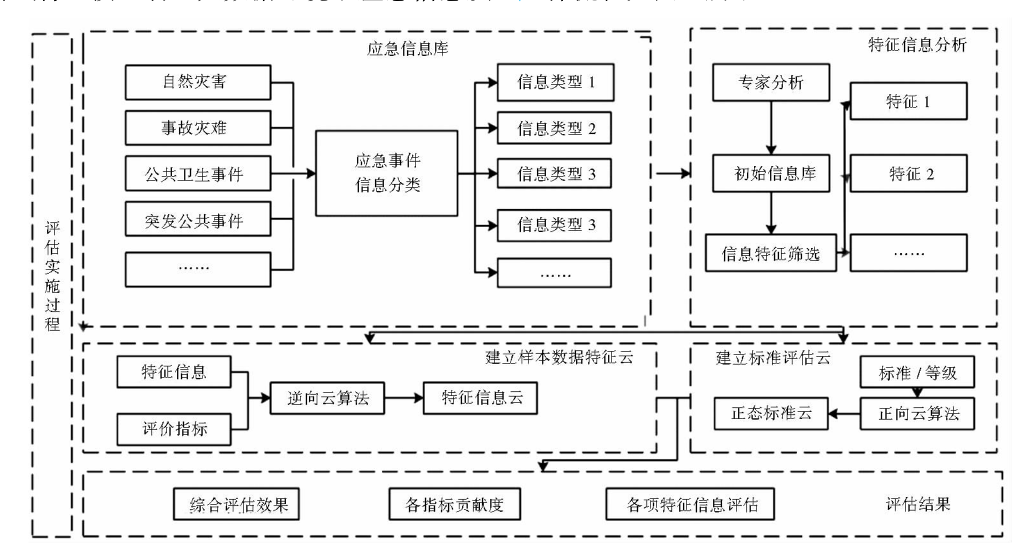


图3 基于云模型的大数据环境下应急信息质量评估流程

4.2 样本选取及数据描述

样本的选取是研究的关键，在对应急信息质量评估时应遵循样本选取的可行性、科学性、代表性等原则，因此笔者在对应急信息数据库现状进行调查研究的基础上，选取了国内具有代表性的EPS数据平台与国外信息较为全面的EM-DAT突发事件数据库，两数据库都具有较为客观全面的应急信息，其应急信息数据对外开放性都优于其他数据库，且应急信息更新较为及时，维护工作完善，在选用国内国外两个应急信息数据库时也考虑了其可对比性。EPS数据平台是国内专业的数据、信息服务提供商，是一个涵盖了经济、工业、农业科技等领域的数据库；EM-DAT突发事件数据库包含了从1900年到现在全世界各类应急信息的重要核心数据，对国内的应急信息收集较客观准确。因此本文随机选取两个数据库中与中国相关的应急信息对两个数据库进行评级，随机选取的内容为中国的疫情信息、病虫灾害信息、森林火灾、干旱、地震、洪水、滑坡、风暴、交通事故共9类应急信息。

4.3 大数据环境下应急信息质量评估案例实施

在对两数据库中的应急信息质量进行评估时，再次邀请前述的应急研究相关领域和应急管理相关部门的15名专家对随机选取的两个数据库中的9类中国应急信息进行打分。

由于专家的研究背景与实践经验的不同，其对每个指标进行评分具有主观性特点，同时打分结果具有模糊性，此外，在同一个环境下同一个评价的专家打分具有一定的随机性，所以可以考虑利用云模型对其进行评估。该项调查的统计数据共有8个指标，专家组根据现实情况访问数据库，并且查阅相关资料，对各指标下的具体情况进行评估，记录包括各项指标的评分结果和得分矩阵，本次评估选用5个评价等级：{非常差，较差，中等，较好，非常好}，即 $S=\{s_{-2}=$“非常差”，$s_{-1}=$“较差”，$s_0=$“中等”，$s_{+1}=$“较好”，$s_{+2}=$“非常好”。

4.3.1 评估指标集

假设S表示目标层，即大数据环境下应急信息质量评估的综合质量。C 表示大数据环境下3个质量维度，C_i 的集合可以用 $C=\{C_1, C_2, \cdots\cdots, C_m\}$ 表示，其含义为信息内容质量、信息描述质量和信息约束质量。C_{ij} 表示 C_i 的二级评估指标的集合，可以表示为 $C_i=\{C_{i1}, C_{i2}, \cdots, C_{ij}\}$，其中 m＝1,2,3，表示3个一级指标，

C_{ij}表示第 i 个一级评估指标中的第 j 个二级指标。

4.3.2 将不确定语言形式的信息转换为云

首先，采用黄金分割法将专家评语集转换为数字特征(z_{-2},…,z_{+2})[29]，假定有效论域为[x_{min},x_{max}] = [0,100]，He_0 = 0.1，则 5 个云依次为：Z_{-2}(0,10.31,0.26)，Z_{-1}(30.9,6.37,0.16)，Z_0(50,3.93,0.1)，Z_{+1}(69.1,3.93,0.16)，Z_{+2}(100,10.31,0.26)。

以(z_{-2},……,z_{+2})为数字特征设定评估等级标准云，可以得出图 4：

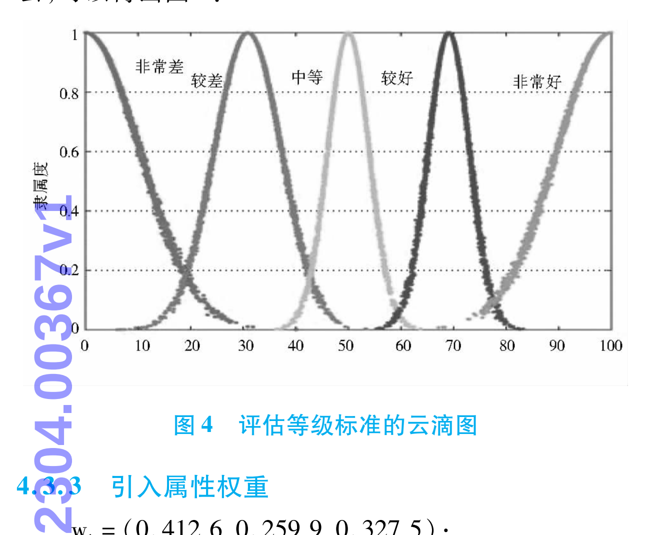


图 4 评估等级标准的云滴图

4.3.3 引入属性权重

w_i = (0.412 6,0.259 9,0.327 5)；

w_{ij} = (0.170 2,0.135 1,0.107 2,0.104,0.052,0.103 9,0.163 8,0.163 8)；

其中(i = 1,2,3；j = 1,2,3,…,8)

4.3.4 构建不同专家下各个案例的多层次综合云

根据 15 位专家对两个数据库对 9 类中国应急信息的评估得分的平均值，可以得出关于大数据环境下应急信息的各指标的评估得分矩阵，根据专家打分结果，由逆向云发生算法公式[30]可得到 A、B 数据库应急信息质量的各项评估指标的云模型数字特征值 C_i(Ex_i,En_j,He_i)(i = 1,2,…,8)

A、B 数据库的 3 个一级指标的计算结果如表 5 所示：

表 5 数据库应急信息质量的一级评估指标云模型数字特征

		C_1	C_2	C_3
A	Ex_i	78.303 6	79.800 7	74.666 7
	En_i	7.196 2	3.554 0	4.697 3
	He_i	2.652 5	1.180 3	1.167 7
B	Ex_i	81.126 8	80.021 7	78.944 4
	En_i	5.746 1	5.649 6	7.935 9
	He_i	2.048 1	2.093 1	3.094 3

A、B 两个数据库 9 个二级指标的结果如表 6 所示：

表 6 数据库中应急信息质量的二级评估指标的云模型数字特征

		C_{11}	C_{12}	C_{13}	C_{21}	C_{22}	C_{23}	C_{31}	C_{32}
A	Ex_i	78.444 4	79.777 8	76.222 2	80.444 4	82.111 1	78.000	74.666 7	74.666 7
	En_i	5.373 9	7.891 2	8.046 0	5.910 7	3.744 5	4.177 7	2.970 8	5.941 6
	He_i	2.494 9	2.706 9	2.820 8	3.105 1	1.382 0	1.340 1	0.821 0	1.432 8
B	Ex_i	81.444 4	81.444 4	80.222 2	78.111 1	81.222 2	81.333 3	77.444 4	80.444 4
	En_i	5.353 7	2.444 7	8.603 0	8.169 8	2.475 7	3.156 5	8.850 6	6.901 0
	He_i	2.148 5	0.548 8	2.903 9	2.575 4	0.752 0	2.009 1	3.248 6	2.931 9

4.3.5

利用云模型对 8 个指标进行加权算数平均计算可以得出样本整体评估的结果：

$CWAA_A$ = (77.501 2,5.896 2,2.153 5)

$CWAA_B$ = (80.124 7,6.522 8,2.450 6)

利用得出的综合指数 $CWAA_A$、$CWAA_B$ 对数字特征与评估等级标准的云滴图进行拟合，可以得出 A、B 信息数据库与标准模型的拟合图(见图 5)。

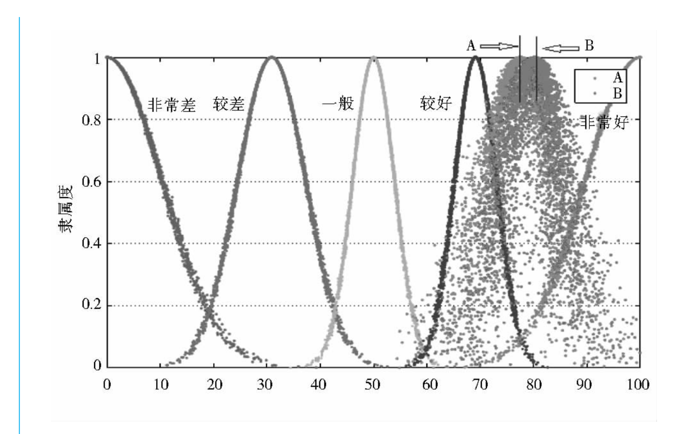


图 5 A、B 应急信息质量与标准云滴图对比

4.4 评估结果分析

(1)从得出的综合评价指数 CWAA 可以分析两个数据库的整体应急信息质量，A、B 两个数据库最能代表应急信息质量的隶属度为 1 的加权平均综合云模型的数字特征值分别为 A = 77.501 2，B = 80.124 7，从而可以初步判断出两个应急信息数据库的信息质量应介于“较好”与“非常好”之间，且更倾向于“较好”，直观上看 B 数据库中的应急信息质量要略优于 A 数据库。

69.1 < A = 77.501 2 < B = 80.124 7 < 100

(2)就各一级指标进行计算分析，在信息内容质

量、信息描述质量、信息约束质量方面，A 数据库质量都稍逊于 B 数据库。但在描述质量维度上，A 数据库的期望特征 Ex_i 要稍小于 B 数据库，而熵特征 En_i 与超熵特征 He_i 却要优于 B 数据库，这就说明在 A 数据库中不同类别的应急信息的描述信息质量差别并不明显，而 B 数据库存在不同类别应急信息的明显差异。就信息约束维度而言，B 数据库质量优于 A 数据库，但是也明显存在不同类型应急信息质量参差不齐的现象。

（3）对比两个应急信息数据库的 8 个二级指标我们可以发现，在 A 数据库中的应急信息的一致性与可衔接性要比 B 数据库好，但在及时性、适用性及有效性方面，B 数据库都要优于 A 数据库。在可获得性、可解释性方面，A 数据库只面向研究机构、政府机构等，需进行注册才可访问，而 B 数据库作为较为全面的国民统计信息数据库，面向对象较为广泛并且应急信息的用户理解性更好。

4.3 评估结果与实际对比

由于缺少对应急信息数据库等级的界定，本文通过将两数据库中的信息从 8 个指标角度与“国家数据”（http://data.stats.gov.cn/）网站中数据进行对比以验证评估结果是否科学。在信息的准确性上 EPS 数据库的准确性基本与“国家数据”中的应急信息基本相同，而 EM-DAT 中的数据存在信息缺失现象，例如在地震观测点、地震灾后恢复重建的支出方面 EM-DAT 数据库并未收录，并且存在有些数据难以验证准确性的情况。在信息的及时性方面，EM-DAT 数据库的地震、风暴信息的及时性要优于 EPS 数据库，而 EPS 的其他几类信息要优于 EM-DAT。在适用性上，EPS 数据库收录的信息为国家各部门对数据的汇总，目的明确，对应急工作有一定指导，而 EM-DAT 数据库中的数据来源于国外相关研究机构、监测机构，目的多为研究使用，因此适用性较差。在一致性、可衔接性上，EM-DAT 中数据的标准、结构、收录数据项目、应急信息编号等都较完善，而且自建库（创建于 1988 年）至今一直沿用，而 EPS 数据库（2005 年）收录年限以及编号等有所欠缺。在信息的解释和可获得性方面，EPS 面向用户为国内机构与居民，而 EM-DAT 面向研究者和国外机构，要稍逊于 EPS。因此评估的结果与实际较为相符。

5 结语与展望

本文研究应急信息质量所处的大数据环境，分析应急管理对于提升信息质量的评估需求，为后期大数据环境下应急信息质量评估体系构建制定需求分析；并且基于理论基础与实践需求，通过德尔菲法构建大数据环境下应急信息质量评估体系，采用层次分析法为指标赋权，该评估体系能较好地反映大数据环境下对信息质量的要求，较为全面地准确反映应急信息质量特征，对指导应急信息质量评估工作具有一定意义；而后利用所构建的大数据环境下应急信息质量评估体系并且结合云模型评估方法对 EPS 数据平台和 EM-DAT 突发事件数据库进行应急信息质量评估，并且将评估结果和实际相对比，分析评估的合理性。

本文研究仍存在一定局限性：应急信息质量衡量标准存在主观性，难以实现自动评估；应急信息的来源过于广泛，但有些难以获取，难以对元数据进行评估。基于此，在大数据环境下应急信息质量评估的研究与实践中要加强相关质量标准的建设。在对元数据进行管理时要实现其标准化，可以通过构建关键业务元数据、及时跟进与理解大数据技术中的元数据、将业务数据与元数据对接以实现对数据追踪、更新应急管理部门的元数据等措施来增加大数据内容。

参考文献：

[1] 查先进，陈明红. 信息资源质量评估研究[J]. 中国图书馆学报，2010，36(2)：46－55.
[2] 宋维翔，贾佳. 微信公众号信息质量与用户互动行为关系研究[J]. 现代情报，2019,39(1):78－85.
[3] 王仙雅，毛文娟，李晋. 信息质量、感知有用性与持续搜寻的关系——基于网络食品安全信息的调查[J]. 情报杂志，2017,36(2):159－164.
[4] 王素芳，白晋铭，黄晨. 高校图书馆信息共享空间服务质量评估研究——以浙江大学为例[J]. 大学图书馆学报，2017,35(2):26－38.
[5] 朱益平，刘春年. 应急信息的数据准确性测量框架研究[J]. 图书馆学研究，2017(13)：88－92,58.
[6] 刘春年，陈通. 基于应急事件的信息服务质量评价实证研究——以应急网站信息服务为例[J]. 情报资料工作，2015(6):68－72.
[7] 袁维海. 应急信息质量模糊评价研究[J]. 华东经济管理，2016,30(4):169－172.
[8] 刘春年,刘宇庆. 基于 PIM 理论的应急专业人员应急产品信息搜索行为研究[J]. 图书馆学研究,2016(17)：40－46,67.
[9] 郭路生，刘春年，胡佳琪. 工程化思维下情报需求开发范式——情报需求工程探析[J]. 情报理论与实践，2017,40(9):24－28.
[10] 郭路生，刘春年. 大数据环境下基于 EA 的政府应急信息资源规划研究[J]. 情报杂志，2016,35(6):171－176.
[11] 郭春侠,徐青梅,储节旺. 大数据时代突发事件应急管理情报分析人才培养初探[J]. 图书情报工作,2019,63(5):14－22.
[12] 李露. 大数据背景下应急管理中的情报信息融合[J]. 信息化

建设, 2016(6):390.
[13] SEPPÄNEN H, VIRRANTAUS K. Shared situational awareness and information quality in disaster management[J]. Safety science, 2015,77(8):112-122.
[14] OLIVEIRA A C M, BOTEGA L C, SARANA J F, et al. Crowdsourcing, data and information fusion and situation awareness for emergency management of forest fires: the project DF100Fogo (FDWithoutFire)[J]. Computers, environment and urban systems, 2019,77(9): 1-9.
[15] STREET R B, BUONTEMPO C, MYSIAK J, et al. How could climate services support disaster risk reduction in the 21st century[J]. International journal of disaster risk reduction, 2019,34(3): 28-33.
[16] YANG W L, STRONG D M, KAHN B K, et al. AIMQ: a methodology for information quality assessment[J]. Information & management, 2003,40(2):133-146.
[17] MADNICK S, ZHU H. Improving data quality through effective use of data semantics[J]. Data & knowledge engineering, 2006,59(2):460-475.
[18] JURAN J M. Juran on planning for quality[M]. New York: Collier Macmillan, 1988.
[19] KAHN B K, STRONG D M, WANG R Y. Information quality benchmarks: product and service performance[J]. Communications of the ACM ,2002,45(4):184-192.
[20] 刘春年, 张曼. 应急信息资源分类目录比较分析及其修订研究[J]. 图书馆学研究, 2014(24): 13-19.
[21] 刘春年, 张凌宇. 基于应急事件的信息资源再生:关键要素与现实路径[J]. 情报理论与实践, 2017, 40(4):66-71.
[22] 齐春泽, 代文锋. 基于云模型的城市灾害应急能力评价[J]. 统计与决策, 2019(4):41-45.
[23] AGGARWAL A. Data quality evaluation framework to assess the dimensions of 3V's of big data[J]. International journal of emerging technology and advanced engineering, 2017,10(7):503-506.
[24] TALEB I, EL Kassabi H T, SERHANI M A, et al. Big data quality: a quality dimension sevaluation[C]//2016 Intl IEEE conferences of ubiquitous intelligence & computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, Internet of people, and smart world congress. Piscataway:IEEE, 2016: 759-765.
[25] HARYADI A F, HULSTIJN J, WAHYUDI A, et al. Antecedents of big data quality: an empirical examination in financial service organizations[C]//2016 IEEE international conference of big data. Piscataway:IEEE, 2016: 116-121.
[26] OSBORNE J, COLLINS S, RATCLIFFE M, et al. What "ideas-about-science" should be taught in school science? a Delphi study of the expert community[J]. Journal of research in science teaching, 2010(7):692-720.
[27] 吴清烈, 蒋尚华. 预测与决策分析[M]. 南京:东南大学出版社,2004.
[28] 郭路生, 刘春年. 大数据时代应急数据质量治理研究[J]. 情报理论与实践, 2016, 39(11): 101-105.
[29] 李德毅, 刘常昱, 杜鹢, 等. 不确定性人工智能[J]. 软件学报, 2004(11):1583-1594.
[30] 吕辉军,王晔, 王德毅, 等. 逆向云在定性评价中的应用[J]. 计算机学报, 2003,26(8):1009-1013.

作者贡献说明:
徐文强:研究设计与论文撰写;
刘春年:框架确定与论文修改;
周涛:参与论文修改和定稿。

Research on Emergency Information Quality Evaluation System in Big Data Environment

Xu Wenqiang　Liu Chunnian　Zhou Tao
School of Management, Nanchang University, Nanchang 330031

Abstract: [**Purpose/significance**] This paper proposes an emergency information quality evaluation system in the big data environment to help screen out high-quality emergency information in view of the new challenges faced by emergency information quality management in the current big data era. [**Method/process**] Based on the characteristics of emergency information in big data environment and the demand analysis for high-quality emergency information in emergency management, an emergency information quality evaluation system in big data environment was designed and constructed. Furthermore, the cloud model was used to evaluate the domestic emergency information in EPS database and EM-DAT emergency database. [**Result/conclusion**] The evaluation results are in good agreement with the actual situation, which indicates that the evaluation system has certain significance for guiding the quality evaluation of emergency information.

Keywords: big data　emergency information　information quality　evaluation system